

EXTRACT

[0008]

[Means for Solving the Problem]

An input/output control system in a computer system according to the present invention having a check point roll back mechanism which performs a process while periodically check points and which returns the state of the system to a state at the time of the latest picked check point when a fault occurs in the system to re-execute the process and an input/output module which writes and reads data in/from a secondary storage device, includes an invalid block management module which manages an address conversion map which allocates at least one physical block to each of logic blocks constituting the secondary storage device and which allocates, to the logic blocks, a pointer for a physical block having a logic block image at the time of the latest check point and a pointer for a logic block having the latest logic block image and a physical block which do not have valid data on the secondary storage device, and is characterized in that, when the input/output module accepts a write request for the secondary storage device, data is written in a physical block different from the latest check point image with respect to the logic blocks, and the physical block is registered in the address conversion table as the latest logic block image, when the input/output module accepts a read request for the secondary storage device, data is read from a physical block having the latest logic block image with respect to the logic blocks, in a check point process,

with respect to the logic blocks on the address conversion map, the physical block having the latest logic block image is set as a physical block having a check point image for the logic blocks and the latest image, in a recovery operation, with respect to the logic blocks on the address conversion map, a physical block having a logic block at the time of a check point as a physical block having a check point image for the logic blocks and the latest image.

[0009]

The input/output control system is characterized in that times at which data are finally written in the logic blocks are further recorded on the address conversion map, when the input/output module accepts a write request for the secondary storage device, data is written in a physical block different from the latest check point image with respect to the logic blocks, the physical block is registered as the latest logic block image, time at which a write process is performed to the physical block is registered on the address conversion map, when the input/output module accepts a read request from the secondary storage device, data is read from a physical block having the latest logic block image with respect to the logic block, in a check point process, the present time is stored as a check point time stamp, in a recovery operation, with respect to the logic blocks on the address conversion map, when a pointer for the latest physical block having a time stamp newer than the check point time stamp is registered, the latest physical block is given to an invalid block management module

and released, and a pointer for a physical block at the time of the check point is registered on the address conversion map as a pointer for the latest physical block.

[0010]

The input/output control system further includes a difference recording mechanism which records, with respect to a logic block subjected to a write process after the time of the latest check point, a combination of the logic block and a physical block allocated to the logic block before the write process, and is characterized in that, when the input/output module accepts a write request for the secondary storage device, the logic block at this time and mapping on an address conversion map of physical blocks on the difference recording mechanism, a physical block newly secured from the invalid block management module on the address conversion map as a physical block for the logic block, a write process is performed to the physical block, in a check point process, all physical blocks recorded on the difference recording mechanism are released for the invalid block management module, all combinations of logic blocks and physical blocks recorded on the difference recording mechanism are deregistered, in a recovery operation, with respect to all the logic blocks recorded on the difference recording mechanism, physical blocks mapped to the logic blocks by the address conversion map are given to the invalid block management module and released, and the combinations of the logic blocks and the physical blocks registered on the difference recording

mechanism are written back to restore the address conversion map at the time of the latest check point.

[0011]

The input/output control system is characterized in that, when the input/output module accepts a write request for the secondary storage device, if a designated logic block is not registered on the difference recording mechanism, mapping of the logic block at this time and a physical block on the address conversion map on the difference recording mechanism, after a physical block newly secured from the invalid block management module is registered on the address conversion map as a physical block for the logic block, a write process is performed to the physical block, if the designated logic block is registered on the difference recording mechanism, data is written in a physical block mapped to the logic block by the address conversion map, in a check point process, all the physical blocks recorded on the difference recording mechanism are released for the invalid block management module, all combinations of logic blocks and physical blocks recorded on the difference recording mechanism are deregistered, in a recovery operation, with respect to the all logic blocks recorded on the difference recording mechanism, physical blocks mapped to the logic blocks by the address conversion map are given to the invalid block management module and released, and the combinations of the logic blocks and the physical blocks registered on the difference recording mechanism are written back to restore the contents of the

address conversion map at the time of the latest check point.

[0012]

According to the above configuration, a computer system using a check point roll back mechanism, since an input/output operation for the secondary storage device can be carried out, the performance of the system is improved. Furthermore, since the number of redundant physical blocks on the secondary storage device can be reduced, the capacity efficiency of the secondary storage device can also be improved.

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-047891

(43)Date of publication of application : 18.02.2000

(51)Int.Cl.

G06F 11/14

G06F 3/06

G06F 12/00

G06F 12/10

(21)Application number : 11-048453

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 25.02.1999

(72)Inventor : SHIMIZU KUNIYASU

(30)Priority

Priority number : 10150041

Priority date : 29.05.1998

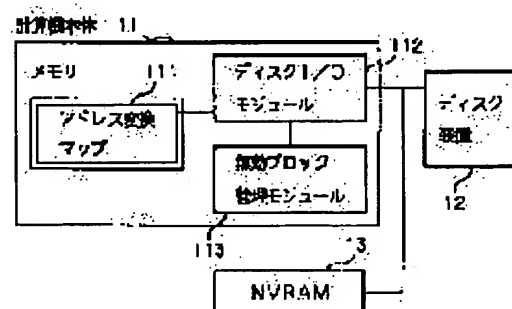
Priority country : JP

(54) DEVICE AND METHOD FOR CONTROLLING INPUT/OUTPUT FOR COMPUTER SYSTEM AND STORAGE MEDIUM STORING THE PROGRAMMED METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To easily realize the capacity efficiency improvement of a secondary storage device.

SOLUTION: This device is provided with an address conversion map 111 assigning at least one logical lock to the logic block of a disk device 12 and assigning a pointer to a physical block with a logic block image at the time point of the newest checking point and a pointer to a physical block with a newest block image to each block, and an ineffective block managing module 113 managing a physical block without effective data on the disk device. Then, in a check point processing, the physical block with the newest logic block image in each logic block on the conversion map is set to be a physical block with a check point image with respect to the logical block and a newest image, and the physical block with the logic block image at the time point of a checking point in each logic block on the conversion map at the time of recovery is set to be a physical block with a check point image with respect to the logic block and the newest image.



LEGAL STATUS

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-47891

(P2000-47891A)

(43) 公開日 平成12年2月18日 (2000.2.18)

(51) Int.Cl. ⁷	識別記号	F I	テームコード* (参考)
G 0 6 F 11/14	3 1 0	G 0 6 F 11/14	3 1 0 B
3/06	3 0 2	3/06	3 0 2 A
12/00	5 0 1	12/00	5 0 1 H
12/10		12/10	E

審査請求 未請求 請求項の数22 O L (全 21 頁)

(21) 出願番号 特願平11-48453

(22) 出願日 平成11年2月25日 (1999.2.25)

(31) 優先権主張番号 特願平10-150041

(32) 優先日 平成10年5月29日 (1998.5.29)

(33) 優先権主張国 日本 (J P)

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 清水 邦保

東京都青梅市末広町2丁目9番地 株式会

社東芝青梅工場内

(74) 代理人 100081732

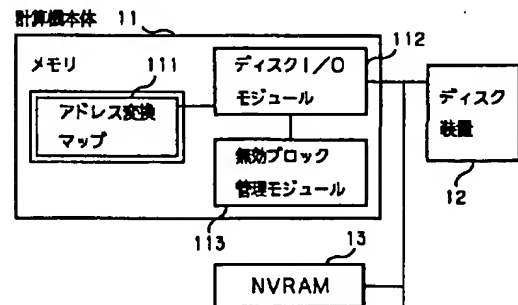
弁理士 大胡 典夫 (外1名)

(54) 【発明の名称】 計算機システムにおける入出力制御装置及び同システムにおける入出力制御方法並びに同方法がプログラムされ記憶された記憶媒体

(57) 【要約】

【課題】 二次記憶装置の容量効率改善を容易に実現すること。

【解決手段】 ディスク装置12の論理ブロックに少なくとも一つの物理ブロックを割当て、各ブロックに最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新論理ブロックイメージを持つ物理ブロックへのポインタを割当てるアドレス変換マップ111、ディスク装置上で有効データを持たない物理ブロックを管理する無効ブロック管理モジュール113を有し、チェックポイント処理では上記変換マップ上の各論理ブロックに最新論理ブロックイメージを持つ物理ブロックを該論理ブロックに対するチェックポイントイメージかつ最新イメージを持つ物理ブロックとし、リカバリ時に上記変換マップ上の各論理ブロックにチェックポイント時点の論理ブロックイメージを持つ物理ブロックを該論理ブロックに対するチェックポイントイメージかつ最新イメージを持つ物理ブロックとする。



【特許請求の範囲】

【請求項1】 定期的にチェックポイントを採用しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、

上記二次記憶装置を構成するそれぞれの論理ブロックに対して少なくとも一つの物理ブロックを割り当て、各論理ブロックに対して、最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ論理ブロックへのポインタを割り当てるアドレス変換マップと、二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして上記アドレス変換マップに登録し、

上記入出力モジュールが上記二次記憶装置に対する読み出し要求を受け付けたときには、その論理ブロックに対する最新の論理ブロックイメージを持つ物理ブロックからデータを読み出し、

チェックポイント処理では、上記アドレス変換マップ上の各論理ブロックに対して、最新の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージかつ最新のイメージを持つ物理ブ

ロックとし、リカバリ時には、上記アドレス変換マップ上の各論理ブロックに対して、チェックポイント時点の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージかつ最新のイメージをもつ物理ブロックとすることを特徴とする計算機システムにおける入出力制御方法。

【請求項2】 定期的にチェックポイントを採用しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、

上記二次記憶装置のそれぞれの論理ブロックに対して一つ以上の物理ブロックを割り当て、各論理ブロックに対して最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当て、更にそれぞれの物理ブロックに最後に書き込み処理を行

なった時刻を記録するアドレス変換マップと、二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして登録すると共にこの物理ブロックに書き込み処理を行なった時刻をアドレス変換マップに登録し、

上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、その論理ブロックに対する最新の論理ブロックイメージを持つ物理ブロックからデータを読み出し、

チェックポイント処理では、現在の時刻をチェックポイントタイムスタンプとして保存し、

リカバリ時には、上記アドレス変換マップ上の各論理ブロックに対して、上記チェックポイントタイムスタンプよりも新しいタイムスタンプを持つ最新の物理ブロックへのポインタが登録されている場合は最新の物理ブロックを上記無効ブロック管理モジュールに渡して解放し、チェックポイント時点の物理ブロックへのポインタを最新の物理ブロックへのポインタとして上記アドレス変換マップに登録することを特徴とする計算機システムにおける入出力制御方法。

【請求項3】 定期的にチェックポイントを採用しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて、

最新チェックポイント時点以降に書き込み処理を行なった論理ブロックに関し、この論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組を記録する差分記録機構と、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、この時点のこの論理ブロックと物理ブロックのアドレス変換マップ上のマッピングを上記差分記録機構に登録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとしてアドレス変換マップに登録し、この物理ブロックに対して書き込み処理を行い、

チェックポイント処理では、上記差分記録機構に登録した全ての物理ブロックを無効ブロック管理モジュールに

解放し、上記差分記録機構に記録してある全ての論理ブロックと物理ブロックの組の登録を抹消し、リカバリ時には、上記差分記録機構に記録してある全ての論理ブロックについて、上記アドレス変換マップでその論理ブロックにマップされている物理ブロックを上記無効ブロック管理モジュールに渡して解放し、上記アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点の上記アドレス変換マップを復元することを特徴とする計算機システムにおける入出力制御方法。

【請求項4】 定期的にチェックポイントを採用しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて、

最新チェックポイント時点以降に書き込み処理を行なった論理ブロックに関し、この論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組を記録する差分記録機構と、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときに、指定された論理ブロックが上記差分記録機構に登録されていなければこの時点のこの論理ブロックと物理ブロックの上記アドレス変換マップ上のマッピングを上記差分記録機構に記録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとしてアドレス変換マップに登録した後、この物理ブロックに対して書き込み処理を行い、指定された論理ブロックが上記差分記録機構に登録されていれば上記アドレス変換マップでこの論理ブロックにマップされている物理ブロックにデータを書き込み、

チェックポイント処理では、上記差分記録機構に登録した全ての物理ブロックを上記無効ブロック管理モジュールに解放し、上記差分記録機構に登録してある全ての論理ブロックと物理ブロックの組の登録を抹消し、

リカバリ時には、上記差分記録機構に登録してある全ての論理ブロックについて、アドレス変換マップでその論理ブロックにマップされている物理ブロックを上記無効ブロック管理モジュールに渡して解放し、上記アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点の上記アドレス変換マップの内容を復元することを特徴とする計算機システムにおける入

出力制御方法。

【請求項5】 定期的にチェックポイントを採用しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、

上記二次記憶装置のそれぞれの論理ブロックに対して一つ以上の物理ブロックを割り当て、各論理ブロックに対して、最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当て、リカバリによって最新チェックポイント時点の状態に戻るメモリ領域および上記二次記憶装置上に配置されるアドレス変換マップと、上記二次記憶装置とメモリの間でアドレス変換マップのページングを行なうアドレス変換マップページング機構と、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとをもち、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレスマップページング機構により上記メモリ上の上記アドレス変換マップの一部の領域を次のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックの最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして上記アドレス変換マップに登録し、

上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレスマップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を次のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対する最新の論理ブロックイメージをもつ物理ブロックからデータを読み出し、

チェックポイント処理では、上記アドレス変換マップ上の各論理ブロックに対して、最新の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージと最新のイメージを持つ物理ブロックとするようにしたことを特徴とする計算機システムにおける入出力制御方法。

【請求項6】 定期的にチェックポイントを採用しながら

ら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、

上記二次記憶装置のそれぞれの論理ブロックに対して一つ以上の物理ブロックを割り当て、各論理ブロックに対して最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当て、リカバリによって最新チェックポイント時点の状態に戻らないメモリ領域および上記二次記憶装置上に配置されるアドレス変換マップと、上記二次記憶装置とメモリの間で上記アドレス変換マップのページングを行なうアドレス変換マップページング機構と、最新チェックポイント時点以降のメモリ上の上記アドレス変換マップの変更履歴を保存するアドレスマップ差分テーブルと、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレスマップページング機構によってメモリ上の上記アドレス変換マップの一部の領域を上記アドレスマップ差分テーブルに登録すると共に、次のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックの最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして上記アドレス変換マップに登録し、

上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合には、上記アドレスマップページング機構により上記メモリ上の上記アドレス変換マップの一部の領域を上記アドレスマップ差分テーブルに登録すると共に、次のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対する最新の論理ブロックイメージを有する物理ブロックからデータを読み出し、

チェックポイント処理では、上記アドレス変換マップ上の各論理ブロックに対して、最新の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージと最新のイメージを持つ物理ブロックとし、上記アドレスマップ差分テーブルに登録され

たエントリを抹消し、

リカバリ時には、上記アドレスマップ差分テーブル登録された各エントリを上記アドレス変換テーブルに書き戻すことを特徴とする計算機システムにおける入出力制御方法。

【請求項7】 定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、

上記二次記憶装置のそれぞれの論理ブロックに対して一つ以上の物理ブロックを割り当て、各論理ブロックに対して最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当て、それぞれの物理ブロックに最後に書き込み処理を行なった時刻を記録し、リカバリによって最新チェックポイント時点に戻るメモリ領域および上記二次記憶装置上に配置されるアドレス変換マップと、上記二次記憶装置とメモリの間で上記アドレス変換マップのページングを行なうアドレス変換マップページング機構と、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップがメモリ上に存在しない場合、上記アドレス変換マップページング機構によりメモリ上の上記アドレス変換マップの一部の領域を次のチェックポイント時点以降に上記二次記憶装置に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとすること、およびこの物理ブロックに書き込み処理を行なった時刻を上記アドレス変換マップに登録し、

上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレスマップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を次のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対する最新の論理ブロックイメージをもつ物理ブロックからデータを読み出し、

チェックポイント処理では、現在の時刻をチェックポイントタイムスタンプとして保存するようにしたこと特徴とする計算機システムにおける入出力制御方法。

【請求項8】 定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、
上記二次記憶装置のそれぞれの論理ブロックに対して一つ以上の物理ブロックを割り当て、各論理ブロックに対して最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当て、それぞれの物理ブロックに最後に書き込み処理を行なった時刻を記録し、リカバリによって最新チェックポイント時点の状態に戻らないメモリ領域および上記二次記憶装置上に配置されるアドレス変換マップと、二次記憶装置とメモリの間でアドレス変換マップのページングを行なうアドレス変換マップページング機構と、最新チェックポイント時点以降のメモリ上のアドレス変換マップの変更履歴を保存するアドレスマップ差分テーブルと、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、
上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレス変換マップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を次のチェックポイント時点以降に上記二次記憶装置に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとすること、およびこの物理ブロックに書き込み処理を行なった時刻を上記アドレス変換マップに登録し、
上記入出力モジュールが二次記憶装置からの読み出し要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレスマップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を上記アドレスマップ差分テーブルに登録すると共に、次のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対する最新の論理ブロックイメージをもつ物理ブロックからデータを読み

出し、

チェックポイント処理では、現在の時刻をチェックポイントタイムスタンプとして保存し、上記アドレスマップ差分テーブルに登録されたエントリを抹消し、
リカバリ時には、上記アドレスマップ差分テーブルに登録された各エントリを上記アドレスマップ差分テーブルに書き戻すようにしたことを特徴とする計算機システムにおける入出力制御方法。

- 【請求項9】 チェックポイント/ロールバック機構を有し、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうリカバリにより最新のチェックポイント時点の状態に戻らないメモリ領域、および上記二次記憶装置上にアドレス変換マップを有し、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて、
最新チェックポイント時点以降に書き込み処理をおこなった論理ブロックに関してこの論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組を記録する差分記録機構と、
20 二次記憶装置とメモリの間で上記アドレス変換マップのページングを行なうアドレス変換マップページング機構と、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、
上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレス変換マップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を次のチェックポイント時点以降に上記二次記憶装置に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、この時点のこの論理ブロックと物理ブロックの上記アドレス変換マップ上のマッピングを上記差分記録機構に記録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとして上記アドレス変換マップに登録し、この物理ブロックに対して書き込み処理を行ない、
40 上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレスマップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を上記アドレスマップ差分テーブルに登録すると共に、
次のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを二次記憶装置から読み込み、その論理ブロックに対する最新の論理ブロックイメージをもつ物理ブロックからデータを読み

出し、
チェックポイント処理では、上記差分記録機構に記録したすべての物理ブロックを上記無効ブロック管理モジュールに解放し、上記差分記録機構に記録してあるすべての論理ブロックと物理ブロックの組の登録を抹消し、リカバリ時には、上記差分記録機構に記録してあるすべての論理ブロックについて、上記アドレス変換マップでその論理ブロックにマップされている物理ブロックを上記無効ブロック管理モジュールに渡して解放し、上記アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点の上記アドレス変換マップを復元するようにしたことを特徴とする計算機システムにおける入出力制御方法。

【請求項 10】 チェックポイント/ロールバック機構を有し、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうリカバリによりて最新チェックポイント時点の状態に戻らないメモリ領域および上記二次記憶装置上に配置されるアドレス変換マップをもち、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なうような入出力モジュールをもつ計算機システムにおいて、

上記二次記憶装置とメモリの間で上記アドレス変換マップのページングを行なう上記アドレス変換マップページング機構をもち、最新チェックポイント時点以降に書き込み処理を行なった論理ブロックに関して、この論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組を記録する差分記録機構と、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールとを有し、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップが上記メモリ上に存在しない場合、上記アドレス変換マップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を次のチェックポイント時点以降に上記二次記憶装置に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、指定された論理ブロックが上記差分記録機構に登録されていなければこの時点のこの論理ブロックと物理ブロックの上記アドレス変換マップ上のマッピングを上記差分記録機構に登録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとして上記アドレス変換マップに登録した後、この物理ブロックに対して書き込み処理を行ない、指定された論理ブロックが上記差分記録機構に登録されていれば上記アドレス変換マップでこの論理ブロックにマップされている物理ブロックにデータを書き込み、

上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、該当する論理ブロックに対応する上記アドレス変換マップがメモリ上に存在しない場合、上記アドレス変換マップページング機構によって上記メモリ上の上記アドレス変換マップの一部の領域を上記アドレス変換マップ差分テーブルに登録すると共に次のチェックポイント時点以降に上記二次記憶装置上に書き出した後にこの領域に該当する論理ブロックに対応する上記アドレス変換マップのエントリを上記二次記憶装置から読み込み、その論理ブロックに対する最新の論理ブロックイメージをもつ物理ブロックからデータを読み出し、

チェックポイント処理では、上記差分記録機構に登録したすべての物理ブロックを上記無効ブロック管理モジュールに解放し、上記差分記録機構に登録してあるすべての論理ブロックと物理ブロックの組の登録を抹消し、リカバリ時には、上記差分記録機構に登録してあるすべての論理ブロックについて、上記アドレス変換マップでその論理ブロックにマップされている物理ブロックを上記無効ブロック管理モジュールに渡して解放し、上記アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点の上記アドレス変換マップを復元するようにしたことを特徴とする計算機システムにおける入出力制御方法。

【請求項 11】 定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態に戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールが記憶装置に割り付けられ格納される計算機システムにおいて、

上記二次記憶装置を構成するそれぞれの論理ブロックに対して少なくとも一つの物理ブロックを割り当て、各論理ブロックに対して最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、ならびに最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当てるアドレス変換マップが割り付けられ保持される記憶手段と、

上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理手段と、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして上記アドレス変換マップに登録し、上記入出力モジュールが上記二次記憶装置に対する読み出し要求を受け付けたときには、その論理ブロックに対する最新の論理ブロックイメージを持つ物理ブ

ロックからデータを読み出し、チェックポイント処理では、上記アドレス変換マップ上の各論理ブロックに対して、最新の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージかつ最新のイメージを持つ物理ブロックとし、リカバリ時には、上記アドレス変換マップ上の各論理ブロックに対して、チェックポイント時点の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージかつ最新のイメージをもつ物理ブロックとする入出力制御手段とを具備することを特徴とする計算機システムにおける入出力制御装置。

【請求項 12】 定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて、

上記二次記憶装置のそれぞれの論理ブロックに対して一つ以上の物理ブロックを割り当て、各論理ブロックに対して最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当て、更にそれぞれの物理ブロックに最後に書き込み処理を行なった時刻を記録するアドレス変換マップが割り付けられ保持される記憶手段と、

二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理手段と、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして登録すると共にこの物理ブロックに書き込み処理を行なった時刻をアドレス変換マップに登録し、上記入出力モジュールが上記二次記憶装置からの読み出し要求を受け付けたときには、その論理ブロックに対する最新の論理ブロックイメージを持つ物理ブロックからデータを読み出し、チェックポイント処理では、現在の時刻をチェックポイントタイムスタンプとして保存し、リカバリ時には、アドレス変換マップ上の各論理ブロックに対して、チェックポイントタイムスタンプよりも新しいタイムスタンプを持つ最新の物理ブロックへのポインタが登録されている場合は最新の物理ブロックを上記無効ブロック管理モジュールに渡して解放し、チェックポイント時点の物理ブロックへのポインタを最新の物理ブロックへのポインタとして上記アドレス変換マップに登録する入出力制御手段とを具備することを計算機システムにおける入出力制御装置。

【請求項 13】 定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直

前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて、

最新チェックポイント時点以降に書き込み処理を行なった論理ブロックに関してこの論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組が記録される差分記録機構と、上記二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理手段と、

上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときには、この時点のこの論理ブロックと物理ブロックの上記アドレス変換マップ上のマッピングを上記差分記録機構に登録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとして上記アドレス変換マップに登録し、この物理ブロックに対して書き込み処理を行い、チェックポイント処理では、上記差分記録機構に登録した全ての物理ブロックを上記無効ブロック管理モジュールに解放し、上記差分記録機構に登録してある全ての論理ブロックと物理ブロックの組の登録を抹消し、リカバリ時には、上記差分記録機構に登録してある全ての論理ブロックについて、上記アドレス変換マップでその論理ブロックにマップされている物理ブロックを上記無効ブロック管理モジュールに渡して解放し、上記アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点の上記アドレス変換マップを復元する入出力制御手段とを具備することを特徴とする計算機システムにおける入出力制御装置。

【請求項 14】 定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて、

最新チェックポイント時点以降に書き込み処理を行なった論理ブロックに関してこの論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組が記録される差分記録手段と、二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理機構と、上記入出力モジュールが上記二次記憶装置への書き込み要求を受け付けたときに、指定された論理ブロックが上

記差分記録機構に登録されていなければこの時点のこの論理ブロックと物理ブロックの上記アドレス変換マップ上のマッピングを上記差分記録機構に登録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとして上記アドレス変換マップに登録した後、この物理ブロックに対して書き込み処理を行い、もし指定された論理ブロックが上記差分記録機構に登録されていれば上記アドレス変換マップでこの論理ブロックにマップされている物理ブロックにデータを書き込み、チェックポイント処理では、上記差分記録機構に登録した全ての物理ブロックを上記無効ブロック管理モジュールに解放し、上記差分記録機構に登録してある全ての論理ブロックと物理ブロックの組の登録を抹消し、リカバリ時には、上記差分記録機構に登録してある全ての論理ブロックについて、上記アドレス変換マップでその論理ブロックにマップされている物理ブロックを上記無効ブロック管理機構に渡して解放し、上記アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点の上記アドレス変換マップの内容を復元する入出力制御手段とを具備することを特徴とする計算機システムにおける入出力制御装置。

【請求項15】 定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置を構成するそれぞれの論理ブロックに対して少なくとも1つの物理ブロックを割り当て、最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージを持つ物理ブロックへのポインタを割り当てるアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて用いられ、上記入出力モジュールが書き込み要求を受け取ったとき、その論理アドレスに対して最新の論理ブロックイメージが存在するか否かをチェックさせる機能と、上記アドレス変換マップを参照することにより、指定された論理ブロックに対して最新の論理ブロックイメージが登録されている場合には、最新の論理ブロックイメージとして登録されている物理ブロックに書き込み処理を行なわせる機能と、指定された論理ブロックに対して最新の論理ブロックイメージが割り当てられていない場合には、無効ブロック管理モジュールから新たに物理ブロックを確保して、この物理ブロックを指定された論理ブロックの最新のイメージを持つ有効物理ブロックとして上記アドレス変換マップに登録し、この物理ブロックに対して書き込みを行なわせる機能とがプログラムされ記憶されるコンピュータ読み取り可能な記憶媒体。

【請求項16】 チェックポイント処理において、アドレス変換マップに登録されている全ての最新チェックポイント時点の論理ブロックイメージとなる物理アドレスエントリを前記無効ブロック管理モジュールに渡すことにより解放し、全ての最新チェックポイント時点の論理ブロックイメージとなる物理アドレスエントリを物理ブロックが割り当てられていないことを示す特定の値に設定させる機能と、リカバリ処理において、上記アドレス変換マップを参照することにより最新チェックポイント時点の論理ブロックイメージとなる物理アドレスエントリが上記特定の値でない全ての論理ブロックに対して対応する有効物理ブロックを上記無効ブロック管理モジュールに渡すことで解放し、最新チェックポイント時点の論理ブロックイメージとなる物理ブロックアドレスエントリの値を最新の論理ブロックイメージとなる物理ブロックアドレスエントリに移動し、最新チェックポイント時点の論理ブロックイメージとなる物理ブロックアドレスエントリに対し上記特定の値を設定させる機能と、システムシャットダウン時には、アドレス変換マップの内容を不揮発性記憶領域に保存し、次回ブート時にはアドレス変換マップの内容をシャットダウン完了時点の状態に復元させる機能とがプログラムされ記録されるコンピュータ読み取り可能な請求項15記載の記憶媒体。

【請求項17】 定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置を構成するそれぞれの論理ブロックに対して少なくとも1つの物理ブロックを割り当て、最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージを持つ物理ブロックへのポインタ、タイムスタンプを割り当てるアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて用いられ、

上記入出力モジュールが書き込み要求を受け取ったとき、最新の論理ブロックイメージとなるブロックのタイムスタンプが最新のチェックポイントタイムスタンプより古いかな否かをチェックさせる機能と、上記アドレス変換マップの内容を参照することにより、指定された論理ブロックに対して有効な物理ブロックのタイムスタンプがチェックポイントタイムスタンプよりも古くない場合には、現在有効な物理ブロックとして登録されている物理ブロックに書き込みを行なわせる機能と、指定された論理ブロックに対して最新の論理ブロックイメージである物理ブロックのタイムスタンプがチェックポイントタイムスタンプよりも古い場合、更に最新チェックポイント時点の論理ブロックイメージであるアドレスが存在するか否かをチェックさせる機能と、指定された論理ブロッ

クの最新チェックポイント時点の論理ブロックイメージである物理ブロックアドレスが上記特定の値でなければ、この物理ブロックを無効ブロック管理モジュールに渡すことで解放し、指定された論理ブロックに対して、最新の論理ブロックイメージである現在有効な物理ブロックのアドレスを最新チェックポイント時点の論理ブロックイメージの物理ブロックアドレスとして登録させる機能と、上記無効ブロック管理モジュールから新たに物理ブロックを確保して、この物理ブロックを指定された論理ブロックの有効物理ブロックとして上記アドレス変換マップに登録し、この物理ブロックに対して書き込みを行なわせる機能とがプログラムされ記憶されるコンピュータ読み取り可能な記憶媒体。

【請求項18】 チェックポイント処理において、現在の時刻をチェックポイントタイムスタンプとして登録させる機能と、リカバリ処理において、アドレス変換マップ上の全ての論理ブロックに対して、その論理ブロックに対応する有効物理ブロックのタイムスタンプがチェックポイントタイムスタンプよりも新しい場合は、有効物理ブロックを上記無効ブロック管理モジュールに渡すことで解放し、最新チェックポイント時点の論理ブロックイメージである物理ブロックアドレスエントリの値を最新の論理ブロックイメージである物理ブロックアドレスに移動し、最新チェックポイント時点の論理ブロックイメージである物理ブロックアドレスエントリの値を上記特定の値に設定させる機能と、システムシャットダウン時には、上記アドレス変換マップを不揮発性記憶に保存することにより、次回ブート時には上記アドレス変換マップをシャットダウン完了時点の状態に復元させる機能とがプログラムされ記憶される請求項15記載のコンピュータ読み取り可能な記憶媒体。

【請求項19】 定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて用いられ、

上記入出力モジュールが書き込み要求を受け取ったとき、指定された論理ブロックに関して、現在の論理ブロックと物理ブロックの組を差分登録機構に登録させる機能と、無効ブロック管理モジュールから新たに物理ブロックを確保し、この物理ブロックを指定された論理ブロックに対するマッピングとして上記アドレス変換マップに登録させる機能と、登録を終えた後この物理ブロックに対して書き込みを行なわせる機能とがプログラムされ記憶されるコンピュータ読み取り可能な記憶媒体。

【請求項20】 チェックポイント処理において、上記

差分登録機構に登録されている全ての物理ブロックを上記無効ブロック管理モジュールに渡すことにより解放し、上記差分登録機構に登録されている全ての論理ブロックアドレスと物理ブロックアドレスの組の登録を抹消させる機能と、リカバリ処理において、上記差分登録機構に登録されてある全ての論理ブロックに関して、上記アドレス変換マップ上でその論理ブロックにマッピングされている物理ブロックを上記無効ブロック管理モジュールに渡すことにより解放し、上記アドレス変換マップに対して、上記差分登録機構に登録されている論理ブロックと物理ブロックのマッピングを上書きすることにより、最新チェックポイント時点のアドレス変換マップの内容を復元させる機能とがプログラムされ記憶される請求項19記載のコンピュータ読み取り可能な記憶媒体。

【請求項21】 定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点にシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置の論理ブロックと物理ブロックのマッピングを行なうアドレス変換マップと、上記二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールを持つ計算機システムにおいて用いられ、

上記入出力モジュールが書き込み要求を受け取ったとき、指定された論理ブロックが差分登録機構に登録されていなければ、指定された論理ブロックに関して現在の論理ブロックと物理ブロックの組を上記差分登録機構に登録させる機能と、無効ブロック管理モジュールから新たに物理ブロックを確保し、この物理ブロックを指定された論理ブロックに対するマッピングとして上記アドレス変換マップに登録させる機能と、登録を終えた後、この論理ブロックにマッピングされた物理ブロックに対して書き込みを行なわせる機能とがプログラムされ記憶されるコンピュータ読み取り可能な記憶媒体。

【請求項22】 チェックポイント処理において、上記差分登録機構に登録されている全ての物理ブロックを上記無効ブロック管理モジュールに渡すことにより解放し、上記差分登録機構に登録されている全ての論理ブロックアドレスと物理ブロックアドレスの組の登録を抹消させる機能と、リカバリ処理において、上記差分登録機構に登録されてある全ての論理ブロックに関して、上記アドレス変換マップ上でその論理ブロックにマッピングされている物理ブロックを上記無効ブロック管理モジュールに渡すことにより解放し、上記アドレス変換マップに対して、上記差分登録機構に登録している論理ブロックと物理ブロックのマッピングを上書きすることにより、最新チェックポイント時点の上記アドレス変換マップの内容を復元する機能とがプログラムされ記憶される請求項21記載のコンピュータ読み取り可能な記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は計算機システムにおける入出力制御方式、特に高信頼計算機システムに用いて好適な入出力制御装置、及び同システムにおける入出力制御方法並びに同方法がプログラムされ記憶された計算機読取り可能な記憶媒体に関する。

【0002】

【従来の技術】高信頼性計算機システムでは、定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には、直前に採取したチェックポイント時点でシステムの状態を戻し（ロールバック）、処理を再実行することによって、システムの一時故障を回避するようなチェックポイント・ロールバック機構を持つものが考えられていた。このチェックポイント・ロールバック機構を内蔵し、かつ、二次記憶装置との間でデータの格納、読み出しを行うような計算機システムにおいては、障害発生によるロールバックの後でも、主記憶やプロセッサ等の状態とディスク装置の内容等入出力装置との整合性を保証する必要がある。

【0003】このため、通常チェックポイント・ロールバック方式においては、オペレーティングシステム（OS）等によって要求された入出力リクエストは、次のチェックポイントによって発行が確定した後にディスク装置に対して発行するものであった。

【0004】

【発明が解決しようとする課題】しかしながら上述した入出力制御方法においては、少なくとも入出力装置の状態を変更する（書き込む）ような要求に関しては、次のチェックポイント採取処理以降に発行を遅延することになっていた。更に、次のチェックポイント処理時以降にチェックポイント時点まで遅延した要求をまとめて発行するため、チェックポイント・ロールバック方式によって高信頼化をはかる計算機システムの性能を劣化させる一つの原因となっていた。

【0005】これを解消するために、二次記憶装置上で一つの論理ブロックに対して必ず二つの物理ブロックを割り当てることにより、最新チェックポイント時点の論理ブロックのイメージを残したまま、もう一つの物理ブロックに新たに発生した書き込み処理を行なう方式が考えられていた。この方式によれば、チェックポイントを待たずに二次記憶装置に対するライト処理を行なうことができる。しかしながら、この場合、実際の物理ディスク容量の50%のデータ領域が有効なディスク容量として使用できないため、物理ディスクの容量効率が悪いといった問題があった。

【0006】一方、二次記憶装置に対して書き込み要求が発生した際には、二次記憶装置に書き込むデータを一旦不揮発性メモリ上に置き、次のチェックポイントを経た後にこの不揮発性メモリ上のデータを二次記憶装置に書き込み、同じくチェックポイントを待たずに二次記

憶装置に対するライト処理を行なう方法が特願平7-151737号で提案されている。しかしながら、このことを実現するためには、不揮発性メモリを含めその周辺部分に特別なハードウェアが必要であるため、実現が容易ではないといった問題があった。

【0007】そこで本発明は上記事情を考慮してなされたものであり、実現が容易で性能劣化が小さく、かつ二次記憶装置の容量効率を改善することのできる上述不具合を解消した計算機システムにおける入出力制御装置、及び同システムにおける入出力制御方法並びに同方法がプログラムされ記録されたコンピュータ読取り可能な記憶媒体を提供することを目的としている。

【0008】

【課題を解決するための手段】本発明の計算機システムにおける入出力制御方式は、定期的にチェックポイントを採取しながら処理を進め、システムに障害が発生した場合には直前に採取したチェックポイント時点でシステムの状態を戻し、処理を再実行するチェックポイント・ロールバック機構を持つと共に、二次記憶装置との間でデータの書き込みおよび読み出しを行なう入出力モジュールをもつ計算機システムにおいて、上記二次記憶装置を構成するそれぞれの論理ブロックに対して少なくとも一つの物理ブロックを割り当て、各論理ブロックに対して、最新チェックポイント時点の論理ブロックイメージを持つ物理ブロックへのポインタ、最新の論理ブロックイメージをもつ物理ブロックへのポインタを割り当てるアドレス変換マップと、二次記憶装置上で有効なデータを持たない物理ブロックを管理する無効ブロック管理モジュールを有し、入出力モジュールが二次記憶装置への書き込み要求を受け付けたときには、その論理ブロックに対して最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして上記アドレス変換マップに登録し、入出力モジュールが二次記憶装置に対する読み出し要求を受け付けたときには、その論理ブロックに対する最新の論理ブロックイメージを持つ物理ブロックからデータを読み出し、チェックポイント処理では、上記アドレス変換マップ上の各論理ブロックに対して、最新の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージかつ最新のイメージを持つ物理ブロックとし、リカバリ時には、上記アドレス変換マップ上の各論理ブロックに対して、チェックポイント時点の論理ブロックイメージを持つ物理ブロックをこの論理ブロックに対するチェックポイントイメージかつ最新のイメージをもつ物理ブロックとすることを特徴とする。

【0009】また、アドレス変換マップには、更にそれぞれの論理ブロックに最後に書き込みを行った時刻を記録し、上記入出力モジュールが二次記憶装置への書き込み要求を受け付けたときには、その論理ブロックに対し

て最新チェックポイントイメージとは異なる物理ブロックに対してデータを書き込み、この物理ブロックを最新の論理ブロックイメージとして登録すると共にこの物理ブロックに書き込み処理を行なった時刻をアドレス変換マップに登録し、入出力モジュールが二次記憶装置からの読み出し要求を受け付けたときには、その論理ブロックに対する最新の論理ブロックイメージを持つ物理ブロックからデータを読み出し、チェックポイント処理では、現在の時刻をチェックポイントタイムスタンプとして保存し、リカバリ時には、アドレス変換マップ上の各論理ブロックに対して、チェックポイントタイムスタンプよりも新しいタイムスタンプを持つ最新の物理ブロックへのポインタが登録されている場合は最新の物理ブロックを無効ブロック管理モジュールに渡して解放し、チェックポイント時点の物理ブロックへのポインタを最新の物理ブロックへのポインタとしてアドレス変換マップに登録することも特徴とする。

【0010】更に、最新チェックポイント時点以降に書き込み処理を行なった論理ブロックに関して、この論理ブロックと、書き込み処理を行なう前にこの論理ブロックに割り当てられていた物理ブロックの組を記録する差分記録機構を有し、入出力モジュールが二次記憶装置への書き込み要求を受け付けたときには、この時点のこの論理ブロックと物理ブロックのアドレス変換マップ上のマッピングを上記差分記録機構に登録し、上記無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとしてアドレス変換マップに登録し、この物理ブロックに対して書き込み処理を行い、チェックポイント処理では、上記差分記録機構に登録した全ての物理ブロックを無効ブロック管理モジュールに解放し、上記差分記録機構に登録してある全ての論理ブロックと物理ブロックの組の登録を抹消し、リカバリ時には、上記差分記録機構に登録してある全ての論理ブロックについて、アドレス変換マップでその論理ブロックにマップされている物理ブロックを無効ブロック管理モジュールに渡して解放し、アドレス変換マップに対して、上記差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点のアドレス変換マップを復元することも特徴とする。

【0011】また、入出力モジュールが二次記憶装置への書き込み要求を受け付けたときに、もし指定された論理ブロックが差分記録機構に登録されていなければこの時点のこの論理ブロックと物理ブロックのアドレス変換マップ上のマッピングを差分記録機構に登録し、無効ブロック管理モジュールから新たに確保した物理ブロックをこの論理ブロックに対する物理ブロックとしてアドレス変換マップに登録した後、この物理ブロックに対して書き込み処理を行い、もし指定された論理ブロックが差分記録機構に登録されていればアドレス変換マップでこ

の論理ブロックにマップされている物理ブロックにデータを書き込み、チェックポイント処理では、上記差分記録機構に登録した全ての物理ブロックを上記無効ブロック管理モジュールに解放し、差分記録機構に登録してある全ての論理ブロックと物理ブロックの組の登録を抹消し、リカバリ時には、上記差分記録機構に登録してある全ての論理ブロックについて、アドレス変換マップでその論理ブロックにマップされている物理ブロックを無効ブロック管理モジュールに渡して解放し、アドレス変換マップに対して、差分記録機構に登録してある論理ブロックと物理ブロックの組を書き戻すことで最新チェックポイント時点のアドレス変換マップの内容を復元することも特徴とする。

【0012】上記構成によれば、チェックポイント・ロールバック方式の計算機システムで、チェックポイントを待たずに二次記憶装置への入出力を発行することができるため、システムのパフォーマンスが向上する。更に、二次記憶装置上で冗長な物理ブロックの数も減少できるため、二次記憶装置の容量効率も改善できる。

【0013】

【発明の実施の形態】以下に本発明の実施形態を図面を参照して説明する。

（第1実施形態）図1は本発明の第1実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図である。図中、11は計算機本体であり、本発明と関係するところでは、アドレス変換マップ111と、ディスク入出力モジュール（ディスクI/Oモジュール）112と、無効ブロック管理モジュール113がメモリに割り付けられ格納される。12は二次記憶装置となる大容量のディスク装置、13は不揮発性メモリ（NVRAM）である。

【0014】通常、ディスク入出力モジュール112は、ファイルシステムからディスク入出力要求を受け取り、ディスク装置12へ入出力要求を発行する。本実施形態によれば、ディスク入出力モジュール112でディスク装置12のデータブロックの論理アドレスと物理アドレスを管理する。

【0015】アドレス変換マップ111の構造を図2に示す。アドレス変換マップ111は、ディスク装置12の全ての論理ブロックに対して有効と旧有効の二つの物理ブロックアドレスのエントリを持つ。

【0016】ここで、有効物理ブロックとは、この論理ブロックの最新のディスクイメージを持つ物理ブロックのことである。旧有効物理ブロックとは、この論理ブロックの最新チェックポイント時点のディスクイメージを持ち、かつ、最新チェックポイント時点以降に対応する論理ブロックに対して書き込み処理が発生したために別の物理ブロックがこの論理ブロックに対して有効物理ブロックとして割り当てられているような物理ブロックのことである。

【0017】全ての論理ブロックに対して一つの有効な物理ブロックが存在し、アドレス変換マップ111の有効物理ブロックアドレスエントリには、この物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在する論理ブロックでは、旧有効物理アドレスエントリに旧有効な物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在しない論理ブロックでは旧有効物理ブロックエントリにはNULLを記録する。

【0018】図3は、無効ブロック管理モジュール113の内部構造を示す。無効ブロック管理モジュール113は、物理ブロックのフリーリストを管理する。即ち、ディスク装置12上の全ての物理ブロックのうち、どの論理ブロックの有効物理ブロックでも旧有効物理ブロックでもない物理ブロック（無効ブロックアドレスリストB、C、…他）を管理するモジュールである。

【0019】以下に、図4に示すディスク入出力モジュール112が書き込み要求を受け取ったときの動作を示すフローチャートを参照して本実施形態の動作について説明する。

【0020】ディスク入出力モジュール112が書き込み要求を受け取ったとき、ディスク入出力モジュール112は、ステップS41にてアドレス変換マップ111を用いて、論理アドレスに対して旧有効アドレスが存在するか否かチェックする。アドレス変換マップ111で、指定された論理ブロックに対して旧有効な物理ブロックアドレスが登録されている場合には（ステップS41のY）、現在有効な物理ブロックとして登録されている物理ブロックに書き込み処理を行なう（ステップS42）。一方、アドレス変換マップ111で、指定された論理ブロックに対して旧有効な論理ブロックがNULLであった場合には（ステップS41のN）、指定された論理ブロックに対して、現在有効な物理ブロックのアドレスを旧有効な物理ブロックアドレスとして登録（ステップS43）し、無効ブロック管理モジュール113から新たに物理ブロックを確保して、この物理ブロックを指定された論理ブロックの有効物理ブロックとしてアドレス変換マップ111に登録する（ステップS44）。そして、この物理ブロックに対して書き込み処理（ステップS42）を行なう。

【0021】ところで、チェックポイント処理では、アドレス変換マップ111に登録されている全ての旧有効物理ブロックを無効ブロック管理モジュール113に渡すことにより解放し、全ての旧有効物理ブロックエントリの値をNULLにする。また、リカバリ処理では、アドレス変換マップ111において、旧有効物理アドレスエントリの値がNULLでない全ての論理ブロックに対して、有効な物理ブロックを無効ブロック管理モジュール113に渡すことで解放し、旧有効物理ブロックアドレスエントリの値を有効物理ブロックアドレスエントリに移動し、旧有効物理ブロックアドレスエントリの値は

NULLとする。更に、システムシャットダウン時には、アドレス変換マップ111を不揮発性メモリ（NVRAM）13に保存することにより、次のブート時にはアドレス変換マップ111をシャットダウン完了時点の状態で復元するものである。

（第2実施形態）図5は本発明の第2実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図である。図中、21は計算機本体であり、本実施形態と関係するところでは、アドレス変換マップ211と、ディスク入出力モジュール212（ディスクI/Oモジュール）と、無効ブロック管理モジュール213、タイムスタンプ機能を有すタイムスタンプ214がメモリに割り付けられ格納される。22は二次記憶装置となる大容量のディスク装置、23は不揮発性メモリ（NVRAM）である。

【0022】ここで示す実施形態では、ディスク入出力モジュール212にて、ディスク装置22のデータブロックの論理アドレスと物理アドレス、および物理ブロックに書き込み処理を行なった時刻を管理する。

【0023】図6は、本実施形態のディスク入出力モジュール212で用いるアドレス変換マップ211の例を示す図である。アドレス変換マップ211では、ディスク装置22の全ての論理ブロックに対して有効と旧有効の二つの物理ブロックアドレスのエントリを持ち、更に、NULLでない物理ブロックアドレスエントリにはその物理ブロックに最後に書き込み処理を行なった時刻を示すタイムスタンプ215を持つ。

【0024】ここで、有効物理ブロックとは、この論理ブロックの最新のディスクイメージをもつ物理ブロックのことである。旧有効物理ブロックとは、この論理ブロックの最新チェックポイント時点のディスクイメージをもち、かつ、最新チェックポイント時点以降に対応する論理ブロックに対して書き込み処理が発生したために別の物理ブロックがこの論理ブロックに対して有効物理ブロックとして割り当てられているような物理ブロックのことである。

【0025】全ての論理ブロックに対して一つの有効な物理ブロックが存在し、アドレス変換マップ211の有効物理ブロックアドレスエントリにはこの物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在する論理ブロックでは、旧有効物理アドレスエントリに旧有効な物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在しない論理ブロックでは旧有効物理ブロックエントリにはNULLを記録する。

【0026】図7に無効ブロック管理モジュール213の構造を示す。無効ブロック管理モジュール213は、図1に示す無効ブロック管理モジュール113同様、物理ブロックのフリーリストを管理する。つまり、ディスク装置22上の全ての物理ブロックのうち、どの論理ブロックの有効物理ブロックでも旧有効物理ブロックでも

ない物理ブロックを管理するモジュールである。ここでは更に、無効ブロック管理モジュール213は定期的にはアドレス変換マップ211を走査し、有効物理ブロックアドレスエントリのタイムスタンプ215が最新チェックポイント時点のタイムスタンプ(チェックポイントタイムスタンプ)214よりも古い場合には、この有効物理ブロックと同じ論理ブロックに割り当てられている旧有効物理ブロックを解放し、この旧有効物理ブロックエントリの値をNULLにする。

【0027】以下、図8に示すディスク入出力モジュール212がライトリクエストを受け取ったときの動作のフローチャートを参照して、本実施形態の動作について説明する。

【0028】ディスク入出力モジュール212が書き込み要求を受け取ったときに、ディスク入出力モジュール212は、まず、有効ブロックのタイムスタンプが最新チェックポイントのタイムスタンプ214より古いかな否かをチェックする(ステップS81)。アドレス変換マップ211で、指定された論理ブロックに対して有効な物理ブロックのタイムスタンプ215がチェックポイントタイムスタンプ214よりも古くない場合には(ステップS81のN)、ステップS82の処理に移り、現在有効な物理ブロックとして登録されている物理ブロックに書き込み処理を行なう。アドレス変換マップ211で、指定された論理ブロックに対して有効な物理ブロックのタイムスタンプ215がチェックポイントタイムスタンプ214よりも古い場合には(S81のY)、更にステップS83において旧有効アドレスが存在するかな否かをチェックする。指定された論理ブロックの旧有効物理ブロックアドレスがNULLでなければ(ステップS83のY)、この旧有効物理ブロックを無効ブロック管理モジュール213に渡すことで解放し(ステップS85)、指定された論理ブロックに対して、現在有効な物理ブロックのアドレスを旧有効な物理ブロックアドレスとして登録し(ステップS84)、無効ブロック管理モジュール213から新たに物理ブロックを確保して、この物理ブロックを指定された論理ブロックの有効物理ブロックとしてアドレス変換マップ211に登録する(ステップS86)。そして、この物理ブロックに対して書き込み処理を行なう(ステップS82)。

【0029】ところで、チェックポイント処理では、最新チェックポイントの時刻をチェックポイントタイムスタンプ214として登録する。リカバリ処理では、アドレス変換マップ211上の全ての論理ブロックに対して、もし、その論理ブロックに対応する有効物理ブロックのタイムスタンプ215がチェックポイントタイムスタンプ214よりも新しい場合は、有効物理ブロックを無効ブロック管理モジュール213に渡すことで解放し、旧有効物理ブロックアドレスエントリの値を有効物理ブロックアドレスに移動し、旧有効物理ブロックアド

レスエントリの値をNULLにする。システムシャットダウン時には、アドレス変換マップ211を不揮発性メモリ(NVRAM)23に保存することにより、次のブート時にはアドレス変換マップ211をシャットダウン完了時点の状態に復元するものである。

(第3実施形態)図9は本発明の第3実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図である。図中、31は計算機本体であり、本実施形態と関係するところでは、アドレス変換マップ311と、ディスク入出力モジュール(ディスクI/Oモジュール)312と無効ブロック管理モジュール313、差分記録機構315がメモリに割り付けられ格納される。32は二次記憶装置となる大容量のディスク装置、33は不揮発性メモリ(NVRAM)である。

【0030】本実施形態では、最新チェックポイント以降に新たに書き込み処理を行なった論理ブロックについて、チェックポイント時点の論理ブロックイメージをもつ物理ブロックを、差分記録機構315によって管理するものである。

【0031】図10に、ディスク入出力モジュール312で用いる差分記録機構315を示す。差分記録機構315は、最新チェックポイント以降に新たに書き込み処理を行なったディスク装置32の論理ブロックに関して、論理ブロックアドレスとチェックポイント時点で割り当てられていた物理ブロックのアドレスの組を記録する。

【0032】図11は、無効ブロック管理モジュール313の構造を示す図である。無効ブロック管理モジュール313は、物理ブロックのフリーリストを管理する。つまり、ディスク記憶装置32上の全ての物理ブロックのうち、どの論理ブロックにも割り当てられていない物理ブロックを管理するモジュールである。

【0033】以下、図12のディスク入出力モジュール312が書き込み要求を受け取ったときの動作を示すフローチャートを参照して本実施形態の動作について説明する。

【0034】まず、ディスク入出力モジュール312が書き込み要求を受け取ったときに、ディスク入出力モジュール312は、指定された論理ブロックに関して、現在の論理ブロックと物理ブロックの組を差分記録機構315に登録し(ステップS121)、無効ブロック管理モジュール313から新たに物理ブロックを確保し、この物理ブロックを指定された論理ブロックに対するマッピングとしてアドレス変換マップ311に登録する(ステップS122)。この後に、この物理ブロックに対して書き込み処理(ステップS123:アドレス変換マップ311で指定した物理ブロックにデータライト)を行なう。

【0035】ところで、チェックポイント処理では、差分記録機構315に登録されている全ての物理ブロック

を無効ブロック管理モジュール313に渡すことにより解放し、差分記録機構315に登録されている全ての論理ブロックアドレスと物理ブロックアドレスの組の登録を抹消する。リカバリ処理では、差分記録機構315に登録してある全ての論理ブロックに関して、アドレス変換マップ311上でその論理ブロックにマッピングされている物理ブロックを無効ブロック管理モジュール313に渡すことにより解放し、アドレス変換マップ311に対して、差分記録機構315に登録している論理ブロックと物理ブロックのマッピングを上書きすることにより、最新チェックポイント時点のアドレス変換マップ311を復元する。尚、アドレス変換マップ311、無効ブロック管理モジュール313、差分記録機構315を、ロールバックするメモリ上に構成した場合には、各構造の状態はロールバックによって自動的にチェックポイント時点の状態を復元できるため、ここでのリカバリ処理は必要なくなる。

【0036】また、システムシャットダウン時には、一対一のアドレス変換マップをもつ通常のシステムと同様に、アドレス変換マップ311を不揮発性メモリ33に保存することにより、次のブート時にはアドレス変換マップをシャットダウン完了時点の状態に復元する。

(第4実施形態) 図13は本発明の第4実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図である。図中、41は計算機本体であり、本発明実施形態と関係するところでは、アドレス変換マップ411と、ディスク入出力モジュール(ディスクI/Oモジュール)412と、無効ブロック管理モジュール413と、差分記録機構415がメモリに割り付けられ格納される。42は二次記憶装置となる大容量のディスク装置、43は不揮発性メモリ(NVRAM)である。

【0037】ここで示す実施形態では、最新チェックポイント以降に新たに書き込み処理を行なった論理ブロックについて、チェックポイント時点の論理ブロックイメージをもつ物理ブロックを、差分記録機構415によって管理する。差分記録機構415は、最新チェックポイント以降に新たに書き込み処理を行なったディスク装置42の論理ブロックに関して、論理ブロックアドレスとチェックポイント時点で割り当てられていた物理ブロックのアドレスの組を記録する(図14参照)。

【0038】図15は無効ブロック管理モジュール413の構造を示す。無効ブロック管理モジュール413は、物理ブロックのフリーリストを管理する。つまり、ディスク装置42上の全ての物理ブロックのうち、どの論理ブロックにも割り当てられていない物理ブロックを管理するモジュールである。

【0039】以下、図16のディスク入出力モジュール412が書き込み要求を受け取ったときの動作を示すフローチャートを参照して本実施形態の動作について説明する。

【0040】ディスク入出力モジュール412が書き込み要求を受け取ったときに、ディスク入出力モジュール412は、まず、その論理アドレスが差分記録機構415に存在するか否かをチェックする(ステップS161)。もし、指定された論理ブロックが差分記録機構415に登録されていないならば(S161のN)、指定された論理ブロックに関して、現在の論理ブロックと物理ブロックの組を差分記録機構415に登録し(ステップS163)、無効ブロック管理モジュール413から新たに物理ブロックを確保し、この物理ブロックを指定された論理ブロックに対するマッピングとしてアドレス変換マップ411に登録する(ステップS164)。この後に、この論理ブロックにマッピングされた物理ブロックに対して書き込み処理を行なう(ステップS162)。一方、差分記録機構415に存在するときは(ステップS161のY)、ステップS162に移行してアドレス変換マップ411により指定された物理ブロックに対して書き込み処理を行う。

【0041】ところで、チェックポイント処理では、差分記録機構415に登録されている全ての物理ブロックを無効ブロック管理モジュール413に渡すことにより解放し、差分記録機構415に登録されている全ての論理ブロックアドレスと物理ブロックアドレスの組の登録を抹消する。リカバリ処理では、差分記録機構415に登録してある全ての論理ブロックに関して、アドレス変換マップ411上でその論理ブロックにマップされている物理ブロックを無効ブロック管理モジュール413に渡すことにより解放し、アドレス変換マップ411に対して、差分記録機構415に登録している論理ブロックと物理ブロックのマッピングを上書きすることにより、最新チェックポイント時点のアドレス変換マップ411の内容を復元する。尚、図9に示す実施形態同様、アドレス変換マップ411、無効ブロック管理モジュール413、差分記録機構415をロールバックするメモリ上に構成した場合には、各構造の状態はロールバックによって自動的にチェックポイント時点の状態を復元できるため、リカバリ処理は必要なくなる。システムシャットダウン時には、一対一のアドレス変換マップをもつ通常のシステムと同様に、アドレス変換マップ411を不揮発性メモリ43に保存することにより、次のブート時にはアドレス変換マップ411をシャットダウン完了時点の状態に復元する。このことにより、チェックポイント・ロールバック方式の計算機システムで、二次記憶装置に対する入出力を遅延なしで発行することにより、システムのパフォーマンスが向上し、かつ、二次記憶装置の容量効率も改善する。

(第5実施形態) 図17は、本発明の第5実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図である。図中、51は計算機本体であり、本発明実施形態と関係するところでは、アドレス変換マップ511

と、ディスク入出力モジュール（ディスクI/Oモジュール）512、無効ブロック管理モジュール513、アドレス変換マップページング機構516がメモリに割り付けられ格納される。52は二次記憶装置となる大容量のディスク装置であり、アドレス変換マップ521を持つ。53は不揮発性メモリ（NVRAM）である。

【0042】ここで示す実施形態では、通常、ディスク入出力モジュール512はファイルシステムからディスク入出力リクエストを受け取り、ディスク装置52へ入出力リクエストを発行する。また、ディスク入出力モジュール512にてディスク装置52のデータブロックの論理アドレスと物理アドレスを管理する。

【0043】図18にディスク入出力モジュール512で用いるアドレス変換マップの例を示す。アドレス変換マップ511ではディスク装置52の全ての論理ブロックに対して有効と旧有効の二つの物理ブロックアドレスのエントリを持つ。つまり、アドレス変換マップ511の各エントリのうち、メインメモリ上には有限個のエントリだけが存在し、その他のエントリはディスク装置52上に存在する。

【0044】ここで、アドレス変換マップ511において、有効物理ブロックとは、この論理ブロックの最新のディスクイメージをもつ物理ブロックのことである。旧有効物理ブロックとは、この論理ブロックの最新チェックポイント時点のディスクイメージをもち、かつ、最新チェックポイント時点以降に対応する論理ブロックに対して書き込み処理が発生したために別の物理ブロックがこの論理ブロックに対して有効物理ブロックとして割り当てられているような物理ブロックのことである。全ての論理ブロックに対して一つの有効な物理ブロックが存在し、アドレス変換マップ511の有効物理ブロックアドレスエントリには、この物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在する論理ブロックでは、旧有効物理アドレスエントリに旧有効な物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在しない論理ブロックでは、旧有効物理ブロックエントリにはNULLを記録する。

【0045】図19に無効ブロック管理モジュール513の構造を示す。無効ブロック管理モジュール513は、物理ブロックのフリーリストを管理する。つまり、ディスク装置52上の全ての物理ブロックのうち、どの論理ブロックの有効物理ブロックでも旧有効物理ブロックでもない物理ブロックを管理するモジュールである。

【0046】以下、図20のディスク入出力モジュール512がライトリクエストを受け取ったときの動作を示すフローチャートを参照して本実施形態の動作について説明する。

【0047】ディスク入出力モジュール512が書き込みリクエストを受け取ったときには、ディスク入出力モジュール512は以下の手順でライトリクエストを発行

する。まず、指定された論理ブロックに対応するエントリがメモリ上のアドレス変換マップ511に存在しない場合は（ステップS502のN）、アドレス変換マップページング機構516がアドレス変換マップ511のページングを行い、指定の論理ブロックに対応するアドレス変換マップ511のエントリをメモリ上に置くものである（ステップS508、S510）。指定された論理ブロックに対応するエントリがメモリ上のアドレス変換マップ511に存在し（ステップS502のY）、アドレス変換マップ511にて指定された論理ブロックに対して旧有効な物理ブロックアドレスが登録されている場合には（S504のY）、現在有効な物理ブロックとして登録されている物理ブロックに書き込み処理を行なう（ステップS506）。アドレス変換マップ511で、指定された論理ブロックに対して旧有効な物理ブロックがNULLであった場合には（ステップS504のN）、指定された論理ブロックに対して、現在有効な物理ブロックのアドレスを旧有効な物理ブロックアドレスとして登録し（ステップS512）、無効ブロック管理モジュール513から新たに物理ブロックを確保して、この物理ブロックを指定された論理ブロックの有効物理ブロックとしてアドレス変換マップ511に登録する（ステップS514）。そして、この物理ブロックに対して書き込み処理を行なう（ステップS506）。

【0048】システムシャットダウン時には、アドレス変換マップ511を不揮発性メモリ53に保存することにより、次のブート時にはアドレス変換マップをシャットダウン完了時点の状態に復元するものである。

【0049】ところで、アドレス変換マップページング機構516の機能を述べるに、メモリ上のアドレス変換マップ511はロールバックで最新チェックポイント時点の状態に戻るものである。

【0050】ディスク装置52上のある論理ブロックに対してアドレス変換マップ511を書き換える必要がある場合、もし、その論理ブロックに対するアドレス変換マップ511のエントリがメモリ上に存在するときは、そのエントリの各有効／旧有効の値を適切に書きかえるだけでよいものである。一方、エントリがメモリ上に存在しない、即ちディスク上にページアウトされているときは、適当なアルゴリズム（LRU等）でメモリ上のアドレス変換マップ511の一つ以上のエントリを選択し、この選択したエントリをディスクへページアウトする。このとき、ページアウト先の論理ブロックに対するアドレス変換マップ511のエントリも正しく設定する。斯様な手順にてメモリ上のアドレス変換マップ511の一部を書き換えるものである。

（第6実施形態）図21は、本発明の第6実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図である。図中、61は計算機本体であり、本発明実施形態と関係するところでは、アドレス変換マップ611

と、ディスク入出力モジュール（ディスクI/Oモジュール）612、無効ブロック管理モジュール613、アドレス変換マップページング機構616、アドレス差分テーブル617がメモリに割り付けられ格納される。62は二次記憶装置となる大容量のディスク装置であり、アドレス変換マップ621を持つ。63は不揮発性メモリ（NVRAM）である。

【0051】ここで示す実施形態では、通常、ディスク入出力モジュール612はファイルシステムからディスク入出力リクエストを受け取り、ディスク装置62へ入出力リクエストを発行する。本実施形態では、ディスク入出力モジュール612でディスク装置62のデータブロックの論理アドレスと物理アドレスを管理する。

【0052】図22にディスク入出力モジュール612で用いるアドレス変換マップ611の例を示す。アドレス変換マップ611では、ディスク装置62のすべての論理ブロックに対して有効と旧有効の二つの物理ブロックアドレスのエントリを持つ。アドレス変換マップ611の各エントリのうち、ロールバックによって最新のチェックポイントの状態に戻らない記憶領域（例えばPCIデバイス上のメモリ）上に有限個のエントリだけが存在し、その他のエントリはディスク装置62上に存在する。

【0053】アドレス変換マップ611において、有効物理ブロックとは、この論理ブロックの最新のディスクイメージをもつ物理ブロックのことである。旧有効物理ブロックとは、この論理ブロックの最新チェックポイント時点のディスクイメージをもち、かつ、最新チェックポイント時点以降に対応する論理ブロックに対して書き込み処理が発生したために別の物理ブロックがこの論理ブロックに対して有効物理ブロックとして割り当てられているような物理ブロックのことである。すべての論理ブロックに対して一つの有効な物理ブロックが存在し、アドレス変換マップ611の有効物理ブロックアドレスエントリにはこの物理ブロックのアドレスを格納する。

【0054】旧有効な物理ブロックが存在する論理ブロックでは、旧有効物理アドレスエントリに旧有効な物理ブロックのアドレスを格納する。旧有効な物理ブロックが存在しない論理ブロックでは旧有効物理ブロックエントリにはNULLを記録するものである。

【0055】図23に無効ブロック管理モジュール613の構成を示す。無効ブロック管理モジュール613は、物理ブロックのフリーリストを管理する。つまり、ディスク装置62上のすべての物理ブロックのうち、どの論理ブロックの有効物理ブロックでも旧有効物理ブロックでもない物理ブロックを管理するモジュールである。

【0056】図24にアドレスマップ差分テーブル617を示す。アドレスマップ差分テーブル617は、最新チェックポイント時点以降、メモリ上のアドレス変換マ

ップ611の変更履歴を保存している。メモリ上のアドレス変換マップ611に割り当てられた領域の内容が、アドレス変換マップ611のページングによってディスク装置62に書き出す際、ページアウトされるエントリの内容をアドレスマップ差分テーブル617に保存する。メモリ上のアドレス変換マップ611のエントリの内容をアドレスマップ差分テーブル617に保存し、ディスク装置62に書き出した後に、このメモリ上のアドレス変換マップ611のエントリは無効になる。新たに別のエントリをディスク装置62上のアドレス変換マップ612からメモリ上のアドレス変換マップ611に読み込む際には、メモリ上のアドレス変換マップ611の無効な領域にのみエントリを読み込むことができる。

【0057】以下、図25に示すディスク入出力モジュール612がライトリクエストを受け取ったときの動作を示すフローチャートを参照して、本実施形態の動作を説明する。

【0058】ディスク入出力モジュール612が書き込みリクエストを受け取ったときには、ディスク入出力モジュール612は以下の手順でライトリクエストを発行する。まず、指定された論理ブロックに対応するエントリがメモリ上のアドレス変換マップ611に存在しない場合は（ステップS602のN）、アドレス変換マップページング機構616がアドレス変換マップ611のページングを行ない、指定の論理ブロックに対応するアドレス変換マップ611のエントリをメモリ上に置き（ステップS610、S612）、ステップS604に移行する。指定された論理ブロックに対応するエントリがメモリ上のアドレス変換マップ611に存在し（ステップS602のY）、アドレス変換マップ611にて指定された論理ブロックに対して旧有効な物理ブロックアドレスが登録されている場合には（ステップS604のY）、現在有効な物理ブロックとして登録されている物理ブロックに書き込み処理を行なう（ステップS608）。一方、アドレス変換マップ611で、指定された論理ブロックに対して旧有効な物理ブロックがNULLであった場合には（ステップS604のN）、指定された論理ブロックに対して現在有効な物理ブロックのアドレスを旧有効な物理ブロックアドレスとして登録し（ステップS614）、無効ブロック管理モジュール613から新たに物理ブロックを確保すると共に、この物理ブロックを指定された論理ブロックの有効物理ブロックとしてアドレス変換マップ611に登録し（ステップS614）、この物理ブロックに対して書き込み処理を行なう（ステップS616）。

【0059】チェックポイント処理では、アドレス変換マップ611に登録しているすべての旧有効物理ブロックを無効ブロック管理モジュールに渡すことにより解放し、すべての旧有効物理ブロックエントリの値をNULLにする。また、リカバリ処理では、アドレス変換マッ

ブ611において、旧有効物理アドレスエントリの値がNULLでないすべての論理ブロックに対して、有効な物理ブロックを無効ブロック管理モジュールに渡すことで解放して、旧有効物理ブロックアドレスエントリの値を有効物理ブロックアドレスエントリに移動し、旧有効物理ブロックアドレスエントリの値はNULLとする。システムシャットダウン時には、アドレス変換マップ611を不揮発性メモリ63に保存することにより、次のブート時にはアドレス変換マップ611をシャットダウン完了時点の状態に復元する。

【0060】ところで、アドレス変換マップページング機構616について述べるに、メモリ上のアドレス変換マップ611はロールバックで最新チェックポイント時点の状態に戻らないものである。

【0061】ディスク装置62上のある論理ブロックに対してアドレス変換マップ611を書き換える必要がある場合、もし、その論理ブロックに対するアドレス変換マップのエントリがメモリ上に存在するときは、そのエントリの各有効/旧有効の値を適切に書きかえるだけでよい。

【0062】もし、エントリがメモリ上に存在しない（ディスク上にページアウトされている）ときは、適当なアルゴリズム（LRU等）でメモリ上のアドレス変換マップ611の一つ以上のエントリを選択し、この選択したエントリをメモリ上のアドレスマップ差分テーブル617にコピーする。また、選択したエントリをディスク装置62へページアウトする。このとき、ページアウト先の論理ブロックに対するアドレス変換マップのエントリも正しく設定する。斯様な手順にてメモリ上のアドレス変換マップ611の一部を置きかえるものである。

【0063】チェックポイント処理の一環で、アドレスマップ差分テーブル617をクリアする。リカバリ処理の一環では、アドレスマップ差分テーブル617に記録した各エントリを、アドレス変換マップ611に書き戻すものである。

【0064】尚、上記各実施形態にて述べた手法は、当該手法が計算機にて実行可能となるようプログラムされたコンピュータ読取り可能な記憶媒体として提供することは勿論可能である。

【0065】

【発明の効果】以上詳記したように本発明によれば、チェックポイント・ロールバック方式の計算機システムで、チェックポイントを待たずに二次記憶装置への入出力を発行することができるため、システムのパフォーマンスが向上する。更に、二次記憶装置上で冗長な物理ブロックの数も減少できるため、二次記憶装置の容量効率も改善できる。更にタイムスタンプを用いることにより、チェックポイント処理が軽くなり一層システムのパフォーマンスが向上する。

【0066】また、図1と図5に示す実施形態が論理ブ

ロックと物理ブロックが1対多数の特殊なマッピングに対して適用できるのに対して、図9と図13に示す実施形態は1対1の一般的なマッピングに対して適用でき、チェックポイント・ロールバック方式の計算機システムにて、チェックポイントを待たずに二次記憶装置への入出力を発行することができるため、システムのパフォーマンスが向上する。更に、二次記憶装置上で冗長な物理ブロックの数も減少できるため、二次記憶装置の容量効率も改善できるものである。

10 【0067】また、図13に示す実施形態では図9に示す実施形態に比較してチェックポイント時の処理が軽くなるため、システムのパフォーマンスが更に向上する。

【0068】更に、本発明によれば、チェックポイント/ロールバック方式の計算機システムで、チェックポイントを待たずに二次記憶装置への入出力を発行することができるため、システムのパフォーマンスが向上できる。加えて、二次記憶装置の容量にかかわらずメインメモリ上やのアドレス変換マップで必要な記憶領域が一定であり、メインメモリを圧迫することはない。チェックポイント処理の負荷も軽減される。

20 【図面の簡単な説明】

【図1】本発明の第1実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図。

【図2】同実施形態に係わるアドレス変換マップの構造を示す図。

【図3】同実施形態に係わる無効ブロック管理モジュールの構造を示す図。

【図4】同実施形態に係わるディスク入出力モジュールの動作手順を示すフローチャート。

30 【図5】本発明の第2実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図。

【図6】同実施形態に係わり、アドレス変換マップの構造を示す図。

【図7】同実施形態に係わる無効ブロック管理モジュールの構造を示す図。

【図8】同実施形態に係わるディスク入出力モジュールの動作手順を示すフローチャート。

【図9】本発明の第3実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図。

40 【図10】同実施形態に係わる差分記録機構の構造を示す図。

【図11】同実施形態に係わる無効ブロック管理モジュールの構造を示す図。

【図12】同実施形態に係わるディスク入出力モジュールの動作手順を示すフローチャート。

【図13】本発明の第4実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図。

【図14】同実施形態に係わる差分記録機構の構造を示す図。

50 【図15】同実施形態に係わる無効ブロック管理モジュールの構造を示す図。

ールの構造を示す図。

【図16】同実施形態に係わるディスク入出力モジュールの動作手順を示すフローチャート。

【図17】本発明の第5実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図。

【図18】同実施形態に係わるアドレス変換マップの構造を示す図。

【図19】同実施形態に係わる無効ブロック管理モジュールの構造を示す図。

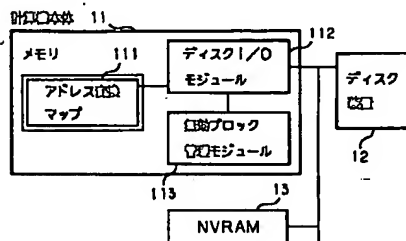
【図20】同実施形態に係わるディスク入出力モジュールの動作手順を示すフローチャート。

【図21】本発明の第6実施形態に係わる高信頼性計算機システムの概略構成を示すブロック図。

【図22】同実施形態に係わるアドレス変換マップの構造を示す図。

【図23】同実施形態に係わる無効ブロック管理モジュールの構造を示す図。

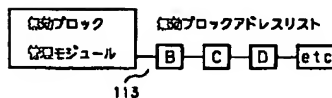
【図1】



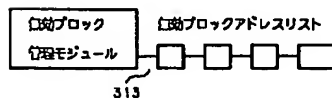
【図2】

物理アドレス	仮想アドレス
0	有効 A
1	有効 B
2	有効 C
3	有効 D

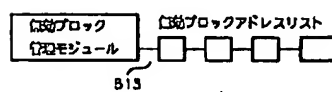
【図3】



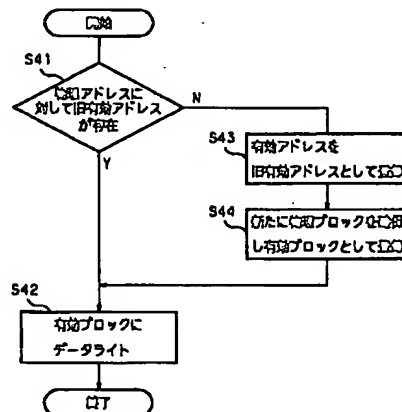
【図11】



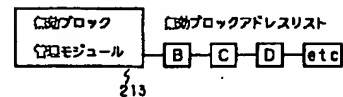
【図19】



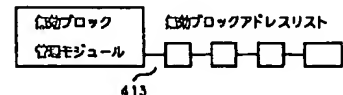
【図4】



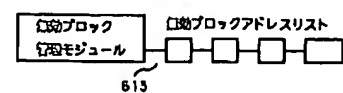
【図7】



【図15】



【図23】



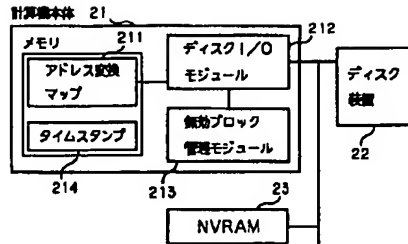
【図24】同実施形態に係わるアドレス差分テーブルの構造を示す図。

【図25】同実施形態に係わるディスク入出力モジュールの動作手順を示すフローチャート。

【符号の説明】

11 (22, 32, 42, 51, 61) …計算機本体、
12 (22, 32, 42, 52, 62) …ディスク装置、
13 (23, 33, 43, 53, 63) …不揮発性メモリ、
111 (211, 311, 411, 511, 611) …アドレス変換マップ、
112 (212, 312, 412, 512, 612) …ディスク入出力モジュール、
113 (213, 313, 413, 513, 613) …無効ブロック管理モジュール、
214 …タイムスタンプ、
315 (415) …差分記録機構、
516 (616) …アドレス変換マップページング機構、
617 …アドレス差分テーブル。

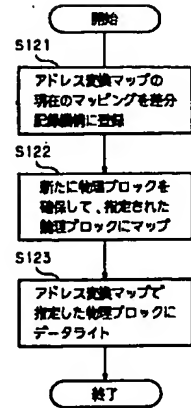
【図5】



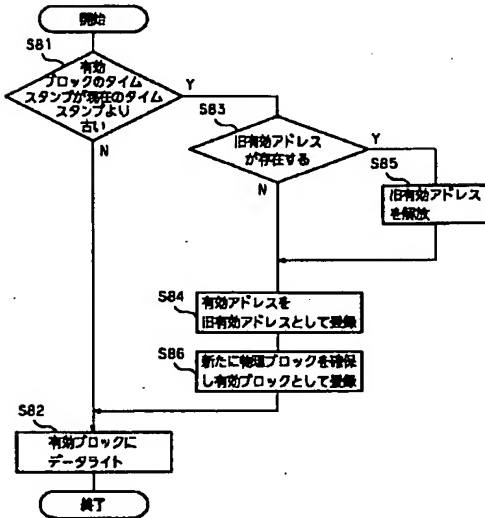
【図6】

論理アドレス		物理アドレス	タイムスタンプ 215
0	有効 旧有効	A a	x t
1	有効 旧有効	B NULL	s -
2	有効 旧有効	C NULL	w -
3	有効 旧有効	D d	v u

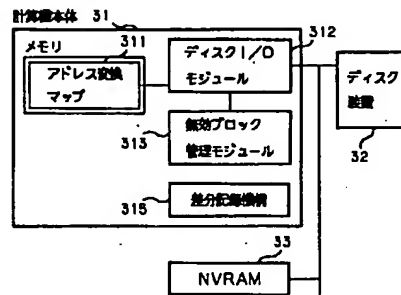
【図12】



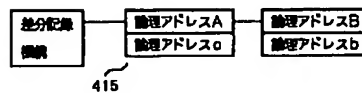
【図8】



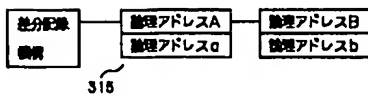
【図9】



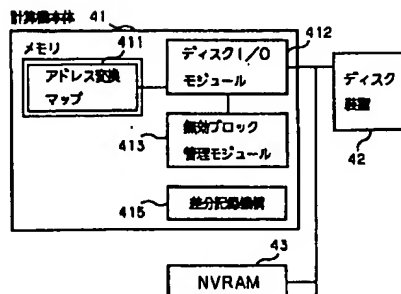
【図14】



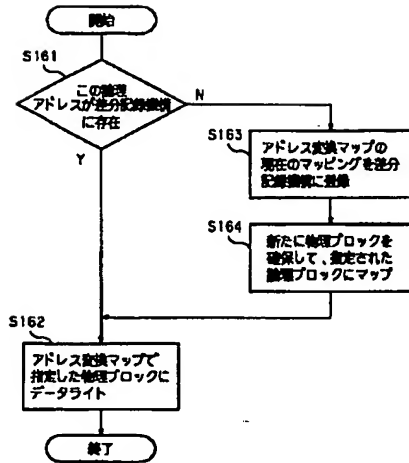
【図10】



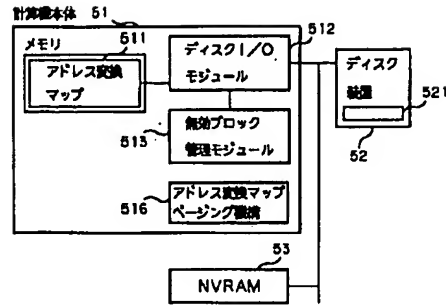
【図13】



【図16】



【図17】

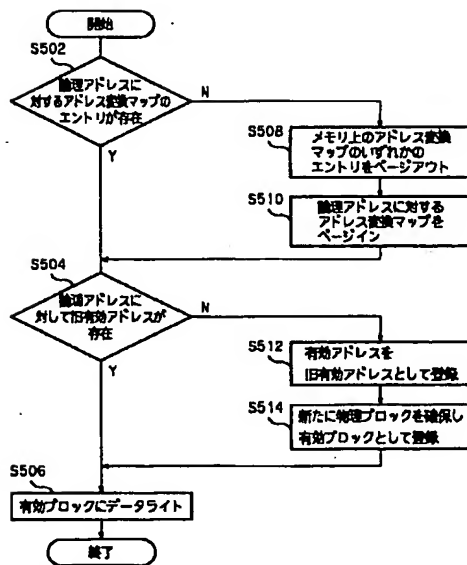


【図18】

論理アドレス		物理アドレス
s	有効	A
	旧有効	a
t	有効	B
	旧有効	NULL
u	有効	C
	旧有効	NULL
v	有効	D
	旧有効	d

511

【図20】

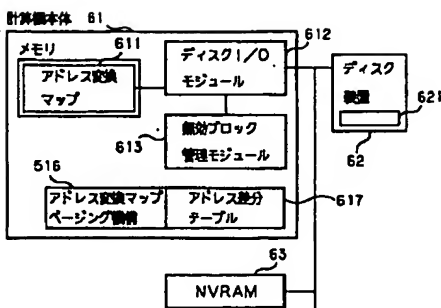


【図22】

論理アドレス		物理アドレス
s	有効	A
	旧有効	a
t	有効	B
	旧有効	NULL
u	有効	C
	旧有効	NULL
v	有効	D
	旧有効	d

611

【図 2 1】



【図 2 4】

論理アドレス		物理アドレス
5	有効	A
	旧有効	a
1	有効	B
	旧有効	NULL
NULL	有効	NULL
	旧有効	NULL
NULL	有効	NULL
	旧有効	NULL

617

【図 2 5】

